American Journal of Epidemiology

Vol. 155, No. 3
Printed in U.S.A.

Rise in Malaria Incidence Rates in South Africa: A Small-Area Spatial Analysis of Variation in Time Trends

Immo Kleinschmidt,[1] Brian Sharp,[1] Ivo Mueller,[2] and Penelope Vounatsou[3]

Using Bayesian statistical models, the authors investigated spatial and temporal variations in small-area malaria incidence rates for the period mid-1986 to mid-1999 for two districts in northern KwaZulu Natal, South Africa. Maps of spatially smoothed incidence rates at different time points and spatially smoothed time trends in incidence gave a visual impression of the highest increase in incidence occurring where incidence rates previously had been lowest. This was confirmed by conditional autoregressive models, which showed that there was a significant negative association between time trends and smoothed baseline incidence before the steady rise in caseloads began. Growth rates also appeared to be higher in the areas close to the Mozambican border. The main findings of this analysis were that: 1) the spatial distribution of the rise in malaria incidence is uneven and strongly suggests a geographic expansion of high-malaria-risk areas; 2) there is evidence of a stabilization of incidence in areas that had the highest rates before the current escalation of rates began; and 3) areas immediately adjoining the Mozambican border appear to have undergone larger increases in incidence, in contrast to the general pattern of low growth in the more northern, high-baseline-incidence areas, but this was not confirmed by modeling. Smoothing of small-area maps of incidence and growth in incidence (trend) is important for interpretation of the spatial distribution of disease incidence and the spatial distribution of rapid changes in disease incidence. *Am J Epidemiol* 2002;155:257–64.

Bayes theorem; epidemiologic methods; incidence; malaria; maps; small-area analysis

The incidence of malaria in South Africa has risen steadily and steeply over the past few years. The total number of cases reported nationally during the first 6 months of 1999 was over 34,000, representing an increase of 80 percent compared with the same period of the previous year (1). Isolated cases of local malaria transmission have been identified in the Durban municipal area, which is several hundred kilometers to the south of malarious areas, and this has given rise to considerable public concern (2). A question that has arisen is whether the increase in caseloads reflects an expansion of South Africa's malaria transmission areas or whether the traditionally malarious areas are suffering increased transmission rates. The sharp increase in malaria incidence could have potentially severe consequences not only for public health but also for tourism and economic development in South Africa.

Received for publication December 8, 2000, and accepted for publication July 9, 2001.

Abbreviations: EPD, expected predictive deviance; SD, standard deviation.

[1] South African Medical Research Council, Durban, South Africa.
[2] Tropical Health Program, University of Queensland, Brisbane, Australia.
[3] Swiss Tropical Institute, Basel, Switzerland.
Reprint requests to Dr. I. Kleinschmidt, South African Medical Research Council, P.O. Box 17120, Congella 4013, South Africa (e-mail: Immo.Kleinschmidt@mrc.ac.za).

The steep increase in malaria in South Africa has been linked to a variety of possible causative factors: the El Niño effect on weather patterns in southern Africa; proximity to areas where no malaria control systems are in place and migration from such areas; the development of drug resistance in the malaria parasite, *Plasmodium falciparum* (3), and insecticide resistance in one malaria vector, *Anopheles funestus* (4); and the possible effects of human immunodeficiency virus infection on a substantial proportion of the South African population (5, 6). In this analysis, we have attempted to document the spatial changes in malaria transmission in two magisterial districts for which high-resolution reporting systems are available. These are the neighboring districts of Ngwavuma and Ubombo in northern KwaZulu Natal, which have hitherto had the highest malaria incidence rates in the country.

The objective of this study was to investigate whether there has been geographic expansion of malaria transmission in South Africa. By modeling the spatial variation of time trends in incidence rates, we sought to establish whether the additional cases are coming predominantly from areas that have always had the highest transmission levels or whether they are originating from previously low-transmission or malaria-free subregions. We used Bayesian statistical methods to produce maps of smoothed incidence at different time points and maps of spatially smoothed rates of change in incidence. By modeling the time trends, we investigated whether there was an association between time trend and baseline malaria incidence.

MATERIALS AND METHODS

In both the Ngwavuma and Ubombo districts, a small-area malaria incidence reporting system has been in operation since 1986 as part of the provincial malaria control program in South Africa (7). This system records all parasitologically confirmed malaria cases, both those found passively and those found by active surveillance. Active surveillance consists of screening measures by which teams of health-care workers go into the community to encourage individuals suspected of having malaria to be tested. Active case-finding forms part of the control strategy of treating all infected individuals. While such active case-finding may not achieve 100 percent coverage, it is thought to identify the vast majority of cases. Since malaria incidence has been generally low in the past (overall annual incidence rates have been below 5 cases per 1,000 in some years), individuals living in the area have, until very recently, had infrequent exposure or no exposure to malaria. The low levels of exposure to *P. falciparum* in the past have made it likely that actively found cases were recent infections rather than cases arising in asymptomatic semi-immune persons. This assumption may no longer be valid in those areas that have the highest incidence rates.

A population census was carried out at the homestead level in 1994 by officials in the malaria control program, which made it possible to use a geographic information system to derive population counts for the same small areas (220 subdivisions known as "sections") for which cases were being reported. Unfortunately, the boundaries for small-area counts from both the 1991 census and the 1996 census are considered unreliable and do not coincide with the section boundaries. Therefore, we used population totals based on the 1994 census, applying a constant and uniform growth rate of 2 percent per annum (8).

The overall malaria incidence rate for the study area over the 13-year period fluctuated strongly around approximately 10 cases per 1,000 person-years annually (see figure 2 below), with neither an upward trend nor a downward trend during the years 1986–1987 to 1994–1995. (We used midyear-to-midyear aggregations, since these correspond roughly to a malaria season.) After 1995, there was a steep and consistent increase in malaria incidence. Therefore, we modeled the data separately for the two time periods 1986–1987 to 1994–1995 and 1995–1996 to 1998–1999.

Modeling

Crude maps of disease incidence are often subject to considerable random error, particularly if either the disease is rare or the population per spatial unit is small, so that the rate may be influenced by a relatively small number of cases. This leads to maps in which attention is drawn to those areas whose rates are based on the least stable estimates (9). Moreover, estimation of the standard errors of explanatory variables will be biased if spatial correlations are not taken into account. These problems can be overcome by spatial smoothing of the rates, which is based on "borrowing strength" from neighboring regions. In this study, we followed the approach that uses hierarchical fully Bayesian spatial modeling as described by Bernadinelli and Montomoli (10). This approach models spatial variation via conditional autoregressive priors (11).

Let Y_{it} and P_{it} denote the observed counts of cases and population, respectively, and let $\eta_{it} \equiv E(Y_{it})$ denote the mean count of cases for the ith area in the tth year. It is assumed that the Y_{it}'s are conditionally independent given the η_{it}'s and that they follow a Poisson distribution, i.e., $Y_{it} \sim \text{Poisson}(\eta_{it})$. The η_{it}'s are defined using customary linear models which may include covariate terms as well as random time and area effects.

The following model was used to estimate smoothed incidence rates for each section (subdivision) in the study area for the 9-year period from 1986–1987 to 1994–1995:

$$\log(\eta_{it}) = \log(P_{it}) + \mu + \varphi_i + \omega_t, \qquad \text{(model 1)}$$

where μ represents the mean incidence rate over all sections over all time periods, φ_i is a random-effects term that allows for spatially structured variation in rates, and ω_t is a random term representing between-year variation, assumed to be independent and normally distributed.

Bayesian statistical inference is based on posterior distributions, which combine information available from the data via the likelihood function and any prior knowledge about the model parameters by specifying appropriate distributions for these parameters. We incorporate our prior information about the structure of the map by assuming conditional autoregressive models for the section (area) random effects. According to the conditional autoregressive model, the section-specific spatial effects φ_i are modeled (conditional on neighboring random effects) as normally distributed with a mean equal to the mean of the effects of its neighbors ($\overline{\varphi_i}$) and a variance that is inversely proportional to the number of neighbors n_i; i.e.,

$$\varphi_i | \varphi_{-i} \sim N(\overline{\varphi_i}, \sigma_\varphi^2 / n_i),$$

where

$$\overline{\varphi_i} = \frac{1}{n_i} \sum_{j \in \text{neighbors of } i} \varphi_j.$$

The effect of this prior distribution is to shrink the incidence rates of sections to the incidence rate of the local mean, where the local mean is the mean of all contiguous sections, excluding the section i itself. The posterior distribution of the rate of a section is therefore a compromise between the prior, which is based on the rates of neighboring sections, and the data for the section, thus stabilizing the rate in sections where the data are sparse due to small populations.

Since no information is available for the remaining parameters, we adopt standard conjugate priors, i.e., vague inverse gamma priors for the variances σ_ω^2 and σ_φ^2 and vague normal priors for all other parameters.

We used a second model to analyze the data for the last 4 years of the series, namely 1995–1996 to 1998–1999. To be able to determine the spatial variation of increases in malaria

incidence, we included in the model a spatially smoothed time trend instead of the random time effects used in model 1.

Bayesian models for the analysis of space-time variation in disease rates have been considered by a number of authors (12–14). Bernadinelli et al. (15) and Sun et al. (16) assumed the temporal variation of the disease rate to be linear, which we judged to be a reasonable constraint in our data given the relatively short period of 4 years for the second time period. We used the following model to estimate the spatially smoothed time trend for each section:

$$\log(\eta_{it}) = \log(P_{it}) + \mu + \varphi_i + (\alpha + \delta_i)t, \qquad \text{(model 2)}$$

where t represents the number of years since 1995–1996, α represents an overall time trend for all sections, and the random term δ_i represents the smoothed local deviation in trend from the overall trend. The latter, which has been termed the *differential trend* (15), is assigned a conditional autoregressive normal prior distribution, as described above, to allow for spatial smoothing of time trends—thereby facilitating the interpretation of patterns in time trend from a map.

A third model was used to investigate the association between time trend and baseline incidence, i.e., average incidence before the period of steady increases in incidence. For this model, the data for the entire time series (13 years) were used, so that baseline incidence (first period) and its effect on differential trend during the second period could be estimated simultaneously from the same model. This method has the advantage that uncertainty in the estimates of baseline incidence are incorporated into the estimates of parameters that express the association between baseline incidence and time trend. The following model was used:

$$\log(\eta_{it}) = \log(P_{it}) + z_{1t}[\mu_1 + \varphi_{1i}] + z_{2t}[\mu_2 + \varphi_{2i} + (\alpha + \delta_i)(t - 9)].$$

The subscripts 1 and 2 refer to the first and second periods of data, respectively, so that μ_1 and μ_2 denote the overall mean logs of incidence rates during the first and second periods, respectively, and φ_{1i} and φ_{2i} denote the section-specific random effects for the first and second periods, respectively. z_{1t} and z_{2t} are indicator variables used to distinguish between the first and second time periods; i.e., $z_{1t} = 1$ for $1 \le t \le 9$ and $z_{1t} = 0$ for $10 \le t \le 13$, while $z_{2t} = 1 - z_{1t}$. The φ_{1i}'s and φ_{2i}'s are assigned separate conditional autoregressive prior distributions, as previously described. The number of years after the start of the second period is represented by $t - 9$. The section-specific differential trend is denoted by δ_i as before. Following the method of Bernadinelli et al. (15), the δ_i's are assumed to be independent, conditional on φ_{1i}, i.e.,

$$[\delta_i|\varphi_{1i}, \sigma_\delta^2] \sim \text{normal}\,(\beta\varphi_{1i}, \sigma_\delta^2). \qquad \text{(model 3)}$$

The term β allows for baseline incidence and trends to be correlated in the prior. It is assigned a vague normal prior. Note that δ_i is modeled conditional on the first-period section effect φ_{1i} to determine its correlation with incidence during the earlier period. The variance term σ_δ^2 represents the variance of the differential trend and is assigned a noninformative gamma distribution. Figure 1 is a graphic representation of model 3, using graphic conventions outlined by Bernadinelli and Montomoli (10) and used in the software package WinBUGS (17).

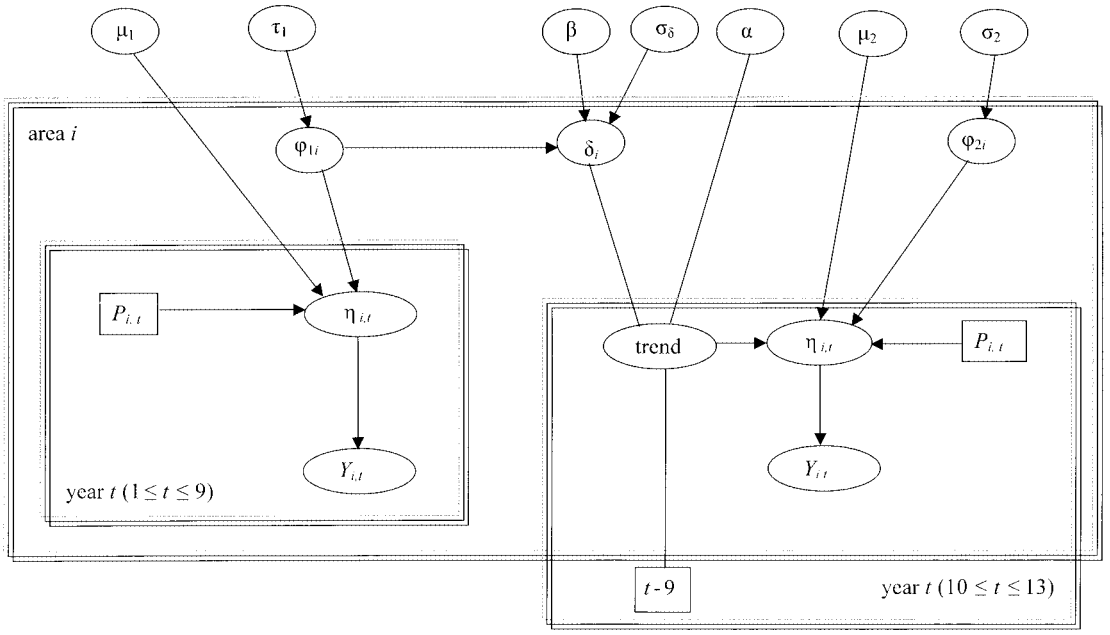

FIGURE 1. Graphic representation of a model (model 3) used to analyze time trends in malaria incidence in two districts (Ngwavuma and Ubombo) in KwaZulu Natal, South Africa, mid-1986 to mid-1999. For definitions of variables, see text.

A further variant of model 3 was used to determine whether sections whose centroids were within 4 km of the Mozambican border followed a different time trend. The investigation of this model was primarily motivated by inspection of the smoothed trend map (see below). A distance of 4 km was chosen because it corresponds approximately to maximum distances of dispersal of the main vector (*Anopheles gambiae*) (18) in the study area. This was done by modeling the differential trend as

$$[\delta_i|\varphi_{1i}, x_i, \sigma_\delta^2] \sim \text{normal}\,(\beta\varphi_{1i} + \gamma x_i, \sigma_\delta^2), \qquad \text{(model 4)}$$

where x_i represents a dummy variable denoting whether a section is within 4 km of the Mozambican border, γ allows the differential trend in border sections to differ from that of other sections, and all other terms have the same meaning as before.

Markov chain Monte Carlo simulation was used to obtain estimates of the posterior and predictive quantities of interest. The models were implemented using Gibbs sampling in the software package WinBUGS (17). To properly monitor convergence, we designed a sampling scheme using three independent chains and a "burn-in" of 12,000 iterations. After convergence, we collected a final sample of 5,000 to obtain summaries of posterior distributions of the parameters. Convergence was assessed using the method of Gelman and Rubin (19).

To compare models 3 and 4, we calculated the expected predictive deviance (EPD) (20) for each model. A brief description of the EPD is given in the Appendix. A lower EPD is indicative of a better model. We also calculated the likelihood ratio statistic, which assesses model fit.

RESULTS

The total population count for the study area was 239,000. Population per section (n = 220) ranged from 11 persons to 5,482 persons (median, 858; mean = 1,086, standard deviation (SD) 884). Area per section ranged from 1.1 km^2 to 263.2 km^2 (median, 19.0; mean = 25.9, SD 27.5). The mean malaria incidence rate per section over the entire study period was 28 cases per 1,000 person-years (SD 91), ranging from an annual mean of 3.4 cases per 1,000 (SD 16) during the year of lowest incidence to an annual mean of 78 cases per 1,000 (SD 133) during the year of highest incidence.

The proportion of very-low-risk sections with incidence rates of less than 1 per 1,000 declined sharply during the period 1995–1996 to 1998–1999 (figure 2). The map of smoothed incidence rates for the period 1986–1987 to 1994–1995, as derived from model 1 (figure 3), clearly shows a trend of the highest incidences' appearing in the northwestern part of the region, with the lowest incidence being seen in the southeast. There are also some high-incidence "hot spots" in the north along the Mozambican border and in the southwest along the Pongola River. The steep rise in incidence rates across the area is evident from the smoothed map of incidence for the final year of the time series (figure 4).

Mean values of spatially smoothed local trend for each section were obtained from model 2. Smoothed local trends for

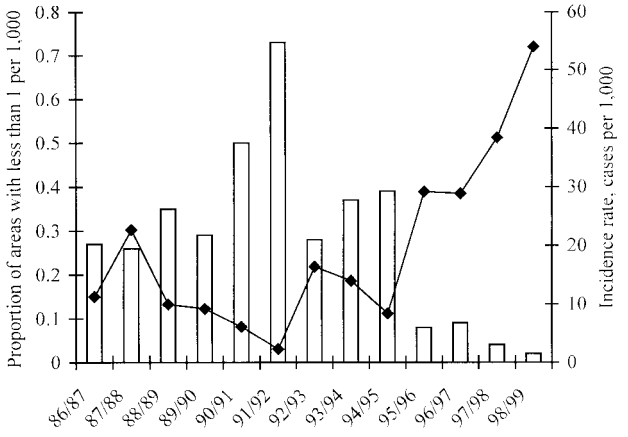

FIGURE 2. Annual malaria incidence rates for two districts (Ngwavuma and Ubombo) and proportion of subdivisions (sections) in those districts with incidence rates of less than 1 per 1,000 person-years, by year, KwaZulu Natal, South Africa, mid-1986 to mid-1999. ◆, annual incidence rate (number of cases per 1,000 person-years); □, proportion of sections with incidence of less than 1 per 1,000.

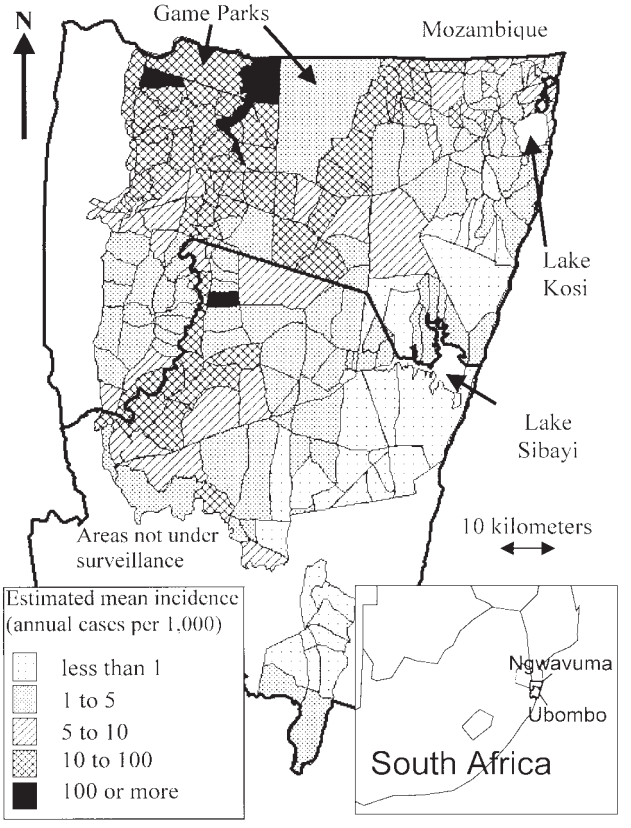

FIGURE 3. Smoothed mean malaria incidence rates, by subdivision (section), estimated from model 1 for two districts (Ngwavuma and Ubombo) in KwaZulu Natal, South Africa, mid-1986 to mid-1995.

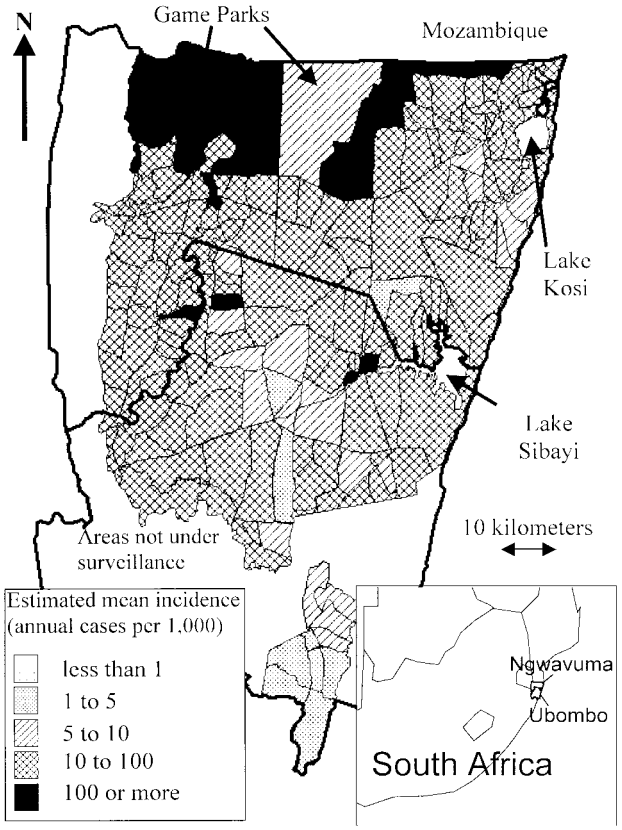

FIGURE 4. Estimated smoothed mean malaria incidence rates, by subdivision (section), for two districts (Ngwavuma and Ubombo) in KwaZulu Natal, South Africa, mid-1998 to mid-1999.

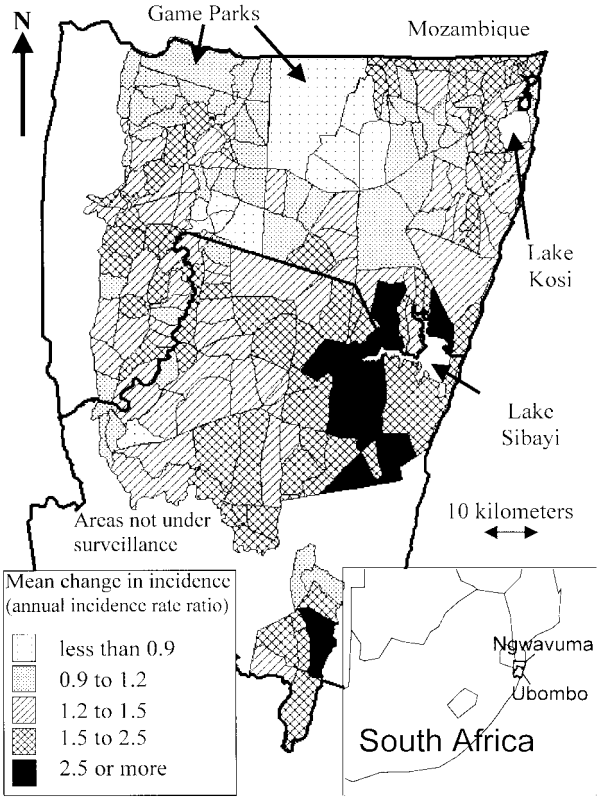

FIGURE 5. Smoothed trend in malaria incidence rates, by subdivision (section), estimated from model 2 for two districts (Ngwavuma and Ubombo) in KwaZulu Natal, South Africa, mid-1995 to mid-1999.

the sections, expressed as incidence rate ratios per annum, varied from 0.6 to 3.5, with a median value of 1.4 and an interquartile range of 1.2–1.7. Figure 5 shows how the smoothed time trends over the 4-year period from 1995–1996 to 1998–1999 were distributed across the study area. This map, when compared with the map of relatively stable "baseline" rates for the period 1986–1987 to 1994–1995 prior to the period of steady growth in annual cases, gives a visual impression of an inverse relation between the gradient of incidence and baseline incidence. The sections with the lowest baseline incidences appear to have been subjected to the steepest increases and vice versa, with some of the high-baseline-incidence sections having experienced either stabilization or a slight negative trend. Sections immediately bordering Mozambique appear to be an exception, with moderately high time trends despite high baseline incidences. The impression of an inverse relation between trend in incidence and baseline incidence is confirmed by figure 6, which is a scatterplot of the log of the trend versus the log of average incidence during the first period.

According to model 3, baseline incidence was significantly negatively associated with trend (tables 1 and 2). According to both model 3 and model 4, incidence rose by just over 40 percent per annum, and this annual rate of increase in incidence was reduced by a factor of approximately 0.90 for each doubling in the baseline rate (incidence in the first period).

According to model 4, the time trend was 16 percent higher in the border sections than in other sections after

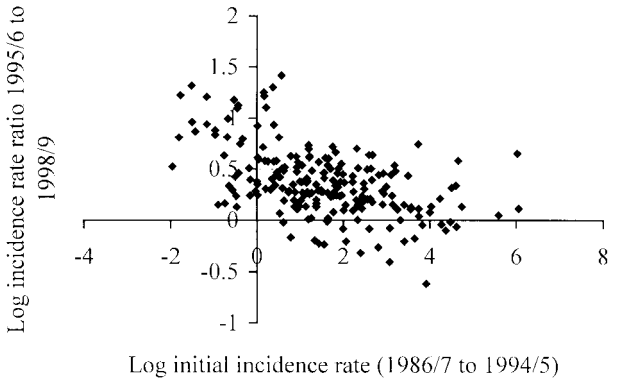

FIGURE 6. Scatterplot of log trend in malaria incidence (1995–1996 to 1998–1999) versus log initial malaria incidence rate (1986–1987 to 1994–1995) for all subdivisions (sections) in two districts (Ngwavuma and Ubombo) in KwaZulu Natal, South Africa.

TABLE 1. Posterior median values for model estimates of malaria incidence (mean log incidence rate, log overall trend, effects of baseline incidence, effects of being a border area, and standard deviation of log differential trend), Ngwavuma and Ubombo districts, South Africa, 1995–1996 to 1998–1999

Parameter symbol	Description of parameter	Estimate			
		Model 3		Model 4	
		Median	95% CI*	Median	95% CI
μ_1	Mean log incidence rate during period 1 (log of cases/person)	−5.42	−5.45, −5.38	−5.42	−5.45, −5.38
μ_2	Mean log incidence rate during period 2 (log of cases/person)	−4.78	−4.82, −4.72	−4.78	−4.83, −4.73
α	Log of overall trend during period 2 (log incidence rate ratio)	0.35	0.31, 0.41	0.34	0.29, 0.38
β	Effect of log baseline incidence on log differential trend (log incidence rate ratio/log incidence rate)	−0.11	−0.14, −0.08	−0.12	−0.14, −0.09
σ_δ	Standard deviation of log differential trend	0.086	0.069, 0.11	0.084	0.067, 0.11
γ	Effect of being a border area on log differential trend (change in log incidence rate ratio)			0.15	0.017, 0.28

* CI, credible interval.

adjustment for the effect of baseline incidence on time trend. However, model comparison (table 2) shows that there is no difference in EPD between model 3 and model 4. Therefore, the data provide no evidence that the differential trend in border sections is different from that in other sections.

DISCUSSION

Our model of smoothed time trends in malaria incidence over the 4 years from 1995–1996 to 1998–1999 showed that there was considerable variation in average annual growth between sections in the districts of Ngwavuma and Ubombo (model 2, figure 5). The average annual increase for all sections was 52 percent. At the two extremes, over 10 percent of the sections ($n = 26$) experienced more than 100 percent annual increases, while just under 10 percent of the sections ($n = 19$) underwent a decline in incidence.

Our map of smoothed trends in incidence rates enabled us to observe a spatial pattern in time trends in relation to baseline incidence. The map of unsmoothed crude trends for each section (not shown) made it difficult to obtain an overall impression because of random noise in the trends. For some sections, it was impossible to calculate a stable crude trend value because of very low incidences or incidences of

TABLE 2. Estimates of model fit criteria and of mean trend, and effects of baseline incidence and proximity to the Mozambican border on annual growth in malaria incidence rates, Ngwavuma and Ubombo districts, South Africa, 1995–1996 to 1998–1999

	Estimate			
	Model 3		Model 4	
	Median	95% CI*	Median	95% CI
Mean trend per annum (e^{α})†	1.42	1.37, 1.50	1.41	1.33, 1.46
Differential trend‡ (effect on annual trend of a doubling of the baseline incidence rate ($e^{\beta \ln 2}$)	0.93	0.91, 0.95	0.92	0.91, 0.94
Effect of being a border area (e^{γ})			1.16	1.02, 1.32
Expected predictive deviance	21,580		21,572	
Likelihood ratio statistic	18,639		18,625	

* CI, credible interval.
† Trend expressed as the incidence rate ratio.
‡ Change in trend expressed as a ratio of incidence rate ratios.

zero at the start of the 4-year period. Smoothing of time trends is important for seeing underlying trends in a map, for the same reasons that smoothing of disease rates to stabilize estimates that are subject to high sampling variability has become commonplace (21).

We were able to confirm the visual impression from the smoothed trend map that increases in incidence rates have been steepest in the sections with the lowest initial rates, using a simple model incorporating initial incidence rates as a factor affecting time trends over the 4 years from 1995–1996 to 1998–1999. The border sections have been of special interest because of their proximity to Mozambique, where spraying of houses with insecticide for malaria vector control is not practiced. The time trend in border sections appeared to be steeper than that in other sections after adjustment for the association of trend with baseline incidence (model 4), but model comparison shows that we do not have evidence against the null hypothesis of border sections' following the same pattern as other sections.

It is therefore evident that for the two districts in our study, there has been a steady geographic expansion of areas at high risk for malaria transmission. It is likely that this expansion has progressed beyond the boundaries of the two districts we analyzed, although this cannot be verified directly, since the malaria reporting system in other districts does not permit the aggregation of cases in small geographic units as is possible for the districts of Ngwavuma and Ubombo.

The negative association of baseline incidence with time trend suggests that there has been some degree of stabilization of rates in high-transmission areas. This could be due to vector- and parasite-related environmental factors. Environmental factors would include the overall climatic suitability for malaria transmission in terms of temperature and rainfall in a region, which may impose constraints on further increases in transmission intensity despite substantial cyclical variation (22–24).

Alternatively, the stabilization of incidence rates in high-transmission areas could be due to population-related factors in the form of a measure of immunity conferred by relatively high levels of exposure. It has been shown in malaria-endemic settings that a small number of infections with *P. falciparum* from birth can lead to an immune response that modulates disease outcome (25). Our own data show that in the high-incidence districts of Ngwavuma and Ubombo, age-specific malaria incidence is lower among adults than it is among teenagers (26). This observation would be consistent with the acquisition of clinical tolerance to infection by at least part of the population in high-incidence subregions of the study area, and it would explain the finding of lower increases in incidence in those regions.

Under circumstances of a steep overall increase in incidence and plausible biologic factors limiting the increase in incidence in some areas, it is unlikely that the observed inverse relation between differential trend and baseline incidence is merely one of regression to the mean. The estimates of baseline incidence in model 3 are fairly precise assessments of each section's underlying incidence during the first period, on account of their being obtained from a long time series of data and their being spatially smoothed. Regression to the mean is unlikely if such stable estimates are used instead of individual unsmoothed observations relating to a particular year.

If immunity rather than leveling of transmission pressures is the cause of stabilization of incidence rates in the high-incidence areas, it is possible that there has been an increase in transmission intensity in excess of that reflected by the incident cases. This would imply that the reported incidence rates are no longer a reliable indicator of transmission in these areas and that the risk to nonimmune visitors may have increased considerably more than the incidence rates. The consequences for tourism and economic development would consequently be more severe.

Our modeling approach represents a modest extension of the spatial-temporal model developed by Bernadinelli et al. (15). We have applied this methodology to an infectious tropical disease with relatively high incidence rates and substantial spatial structure in the data (24). We were able to model the data from the two time periods of the study comprehensively within the same overall model, which enabled us to test the effects of incidence rates in the first period on time trends in the second period. In future work on this topic, we plan to investigate the potential association between possible model covariates (such as rainfall) in an earlier period and malaria incidence in a later period.

The main limitation of this study was its dependence on data obtained from the provincial malaria control program. One concern is that underreporting of cases may be worse in high-incidence areas because resources are overstretched in those areas, and this may have biased our trend analysis. Furthermore, we were unable to incorporate possible between-area population movements into our analysis.

In conclusion, this retrospective analysis found that the spatial distribution of the recent rise in malaria incidence in South Africa is uneven, and strongly suggests an expansion of areas of high malaria risk in South Africa. There is evidence of a stabilization of incidence in areas that had the highest rates before the current escalation of rates began. The impression that subregions immediately adjoining the Mozambican border have undergone larger increases in incidence, in contrast to the general pattern of stabilization of rates in the more northern, high-baseline-incidence areas, was not confirmed by modeling.

Smoothing of small-area maps of incidence and growth in incidence (trend) is important for interpretation of the spatial distribution of disease and rapid changes in disease incidence. Analysis of time trend in relation to baseline incidence provides a useful means of describing geographic expansion of disease risk, provided that precise estimates of baseline incidence can be made.

ACKNOWLEDGMENTS

This work was partially funded by the Swiss National Science Foundation (grant SNF 3200-057165.99) and the MARA/ARMA Project.

The authors thank Drs. Tom Smith and Linda Haines for their valuable comments.

REFERENCES

1. Department of Health, Republic of South Africa. Notifiable medical conditions. Epidemiol Comments 1999;1:11–12.
2. Pillay T. Malaria scare hits Durban. Johannesburg, South Africa: Sunday Times, April 16, 2000:5.
3. Bredenkamp BL, Sharp BL, Mthembu SD, et al. Failure of sulphadoxine-pyrimethamine in treating *Plasmodium falciparum* malaria in KwaZulu-Natal province. S Afr Med J 2001;91: 970–1.
4. Hargreaves K, Koekemoer LL, Brooke BD, et al. *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. Med Vet Entomol 2000;14:181–9.
5. Department of Health, Republic of South Africa. Tenth national HIV survey in women attending antenatal clinics in public health services in South Africa, Oct/Nov 1999. Pretoria, South Africa: Department of Health, 2000.
6. Whitworth J, Morgan D, Quigley M, et al. Effect of HIV-1 and increasing immunosuppression on malaria parasitaemia and clinical episodes in adults in rural Uganda: a cohort study. Lancet 2000;356:1051–6.
7. Sharp B, Fraser C, Naidoo K, et al. Computer-assisted health information system for malaria control. Presented at the MIM [Multilateral Initiative on Malaria] African Malaria Conference, Durban, South Africa, March 14–19, 1999.
8. Statistics South Africa. Mid-year estimates. (Statistical release P0302). Pretoria, South Africa: Statistics South Africa, 1999.
9. Cuzick J, Elliott P. Small-area studies: purpose and methods. In: Cuzick J, Elliott P, eds. Geographical and environmental epidemiology: methods for small-area studies. Oxford, United Kingdom: Oxford University Press, 1992:14–21.
10. Bernadinelli L, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. Stat Med 1992;11:983–1007.
11. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43:671–81.
12. Heisterkamp SH, Doornbos G, Nagelkerke NJ. Assessing health impact of environmental pollution sources using space-time models. Stat Med 2000;19:2569–78.
13. Knorr-Held L, Besag J. Modelling risk from a disease in time and space. Stat Med 1998;17:2045–60.
14. Waller LA, Carlin BP, Xia H, et al. Hierarchical spatio-temporal mapping of disease rates. J Am Stat Assoc 1997;92: 607–17.
15. Bernadinelli L, Clayton D, Pascutto C, et al. Bayesian analysis of space-time variation in disease risk. Stat Med 1995;14: 2433–43.
16. Sun D, Tsutakawa RK, Kim H, et al. Spatio-temporal interaction with disease mapping. Stat Med 2000;19:2015–35.
17. WinBUGS (Bayesian inference Using Gibbs Sampling) software, version 1.2. Cambridge, United Kingdom: MRC Biostatistics Unit, 1999. (MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, United Kingdom). (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml).
18. Gillies MT, De Meillon B. The Anophelinae of Africa south of the Sahara. Johannesburg, South Africa: South African Institute for Medical Research, 1968:213.
19. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci 1992;7:457–72.
20. Carlin BP, Louis TA. Bayes and empirical Bayes methods for data analysis. London, United Kingdom: Chapman and Hall Ltd, 1996.
21. Wakefield JC, Best NG, Waller L. Bayesian approaches to disease mapping. In: Elliott P, Wakefield JC, Best NG, et al, eds. Spatial epidemiology: methods and applications. Oxford, United Kingdom: Oxford University Press, 2000:104–27.
22. Craig MH, Snow RW, le Sueur D. A climate-based distribution model of malaria transmission in Africa. Parasitol Today 1999;15:105–11.
23. Molineaux L. The epidemiology of human malaria as an explanation of its distribution, including some implications for its control. In: Wernsdorfer WH, McGregor I, eds. Malaria: principles and practice of malariology. Vol 2. London, United Kingdom: Churchill Livingstone, 1988.
24. Kleinschmidt I, Sharp BL, Clarke GP, et al. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. Am J Epidemiol 2001;153:1213–21.
25. Gupta S, Snow RW, Donnelly C, et al. Immunity to severe malaria is acquired after one or two exposures. Nat Med 1999; 5:340–3.
26. Kleinschmidt I, Sharp B. Patterns in age-specific malaria incidence in a population exposed to low levels of malaria transmission intensity. Trop Med Int Health 2001;6:986–91.

APPENDIX

Expected Predictive Deviance

We assessed model fit by comparing the expected predictive deviance (EPD) (20) of different models. For the Poisson model, this can be computed as follows: Let l denote all combinations of i,t, and let $y_{l,\text{new}}$ refer to the posterior replicates of the observed data $y_{l,\text{obs}}$. The $y_{l,\text{new}}$ are obtained by sampling Poisson$(E_l \exp(\mu_l))$, with posterior samples of μ_l obtained after convergence of the Markov chain Monte Carlo run. The EPD is calculated as

$$E[d(\boldsymbol{y}_{\text{new}}, \boldsymbol{y}_{\text{obs}})|\boldsymbol{y}_{\text{obs}}, M_i],$$

where

$$d(\boldsymbol{y}_{\text{new}}, \boldsymbol{y}_{\text{obs}}) = 2\Sigma_l \{y_{l,\text{obs}} \log(y_{l,\text{obs}}/y_{l,\text{new}}) - (y_{l,\text{obs}} - y_{l,\text{new}})\}.$$

Carlin and Louis (20, pp. 232–233) show that the EPD consists of two components—namely, a likelihood ratio statistic (LRS) indicating goodness of fit and a penalty term (PEN) which penalizes for under- or overfitting—i.e., EPD = LRS + PEN. Smaller values of EPD are indicative of a better model. The two components can be calculated from

$$\text{LRS} = d(E[y_{\text{new}}|y_{\text{obs}}, M_i], y_{\text{obs}})$$

and

$$\text{PEN} = 2\Sigma_l (y_{l,\text{obs}})\{\log E[y_{l,\text{new}}|y_{\text{obs}}] - E[\log(y_{l,\text{new}})|y_{\text{obs}}]\}.$$

All three terms can be computed from Markov chain Monte Carlo samples.